# 2
# Methods

## 2.1 Habitat Quality Assessment

Habitat quality assessment is an essential part of any assessment of ecological integrity (Karr et al. 1986; Plafkin et al. 1989). The quality of the physical habitat at a site identifies constraints on the attainable biological potential of that site and provides information for interpreting biosurvey results (Barbour and Stribling 1991). Numerous components of the physical structure of stream environments and riparian habitat are critical to the ecological integrity of lotic water resources, including channel morphology (width, depth, and sinuosity); floodplain shape and size; channel gradient; instream cover (boulders, woody debris); substrate type and diversity; riparian vegetation and canopy cover; and bank stability.

Specific habitat parameters and narrative descriptions of the condition categories for which visual assessments of condition are made are shown in Figure 2-1. Some scoring systems have some habitat characteristics weighted more heavily than others. For instance, the parameter condition scoring framework (Barbour and Stribling 1991) used for the 1992 Ohio study had differential weighing for the primary, secondary, and tertiary parameters with a maximum of 20, 15, and 10 points, respectively. However, with the testing of habitat assessment consistency among multiple observers (Barbour and Stribling 1994), it became evident that the weighing could be a substantial source of variability. The habitat scoring systems currently recommended have all parameters weighted equally (Figure 2-1); that is, on a 20-point scale. The scoring system used in New York used equal weighing.

Parameters are visually inspected at each sampling location and assigned scores within the continuum of conditions ranging from optimal to poor based on the narratives. The scores assigned to each parameter are totalled for a station. That score is compared to the reference score to provide a relative assessment of habitat quality that will assist in the interpretation of biological condition. The total score for each sampling station is used in classifying the station, based on the percent comparability to the reference condition ("expected" condition) and the station's apparent potential to support the same level of biological community development as that observed at the reference station. Basic water quality data (temperature, dissolved oxygen, pH, and conductivity) are also collected to allow for further comparison among sites. Further discussion of the logic and justification for the approach can be found in several other documents (Plafkin et al. 1989; Barbour and Stribling 1991, 1994).

## 2.2 Benthic Macroinvertebrate Sampling

For the benthic macroinvertebrate studies, a standardized collection procedure based on RBPs (Plafkin et al. 1989) was used to obtain samples of the macroinvertebrate fauna from comparable habitat types at all stations. Sampling, according to RBPs for high-gradient streams, is focused on what is generally considered to be the most productive of stream systems, riffles and runs. For the New York study, three different RBP level assessments were conducted at each station in order to compare assessment results from the differing levels of effort (RBPI, RBPII, RBPIII).

### 2.2.1 Sampling and Sample Handling

Samples were obtained using a 1-$m^2$ kick net (no. 30 mesh, 600 µm openings). Two 1-$m^2$ samples were collected at each station: one from a fast-water riffle and one from a slow-water riffle. Sampling from both the fast and slow riffle current velocities allows for a broader coverage of variability within the riffle habitat. For those sampling sites which lacked riffles, run areas with cobble or gravel substrate were sampled instead. The two kick net samples from each station were composited in the field, concentrated in a no. 30 (600 µm) sieve bucket, and emptied into a gridded sorting pan for subsampling. For the Ohio portion of the study, the gridded pan was a metal, porcelain-covered pan with numbered grid squares drawn on the bottom. For New York, a change in subsampling methods was made to minimize movement of organisms among grids and increase the standardization of the subsampling effort. The standardized gridded screen (Caton 1991) contains 30 clearly marked squares, each a uniform 6 cm x 6 cm. The gridded screen fits into another slightly larger tray so that water can be added to the sample to allow for even distribution. When the screen is lifted out of the tray, the sample contents settle onto the screen, effectively restricting organism mobility.

**Figure 2-1. Habitat scoring system for streams with riffle/run prevalence.**

HABITAT ASSESSMENT FIELD DATA SHEET *RIFFLE/RUN PREVALENCE*

| Habitat Parameter | Category: Optimal | Suboptimal | Marginal | Poor |
|---|---|---|---|---|
| **1. Instream Cover** | Greater than 50% mix of boulder, cobble, sub-merged logs, undercut banks, or other stable habitat. | 30-50% mix of boulder, cobble, or other stable habitat; adequate habitat. | 10-30% mix of boulder, cobble, or other stable habitat; habitat availability less than desirable. | Less than 10% mix of boulder, cobble, or other stable habitat; lack of habitat is obvious. |
| SCORE ____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **2. Epifaunal Substrate** | Well-developed riffle and run; riffle is as wide as stream and length extends two times the width of stream; abundance of cobble. | Riffle is as wide as stream but length is less than two times width; abundance of cobble; boulders and gravel common. | Run area may be lacking; riffle not as wide as stream and its length is less than 2 times the stream width; gravel or large boulders and bedrock prevalent; some cobble present. | Riffles or run virtually nonexistent; large boulders and bedrock prevalent; cobble lacking. |
| SCORE ____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **3. Embeddedness** | Gravel, cobble, and boulder particles are 0-25% surrounded by fine sediment. | Gravel, cobble, and boulder particles are 25-50% surrounded by fine sediment. | Gravel, cobble, and boulder particles are 50-75% surrounded by fine sediment. | Gravel, cobble, and boulder particles are more than 75% surrounded by fine sediment. |
| SCORE ____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **4. Velocity/Depth Regimes** | All four velocity/ depth regimes present (slow-deep, slow-shallow, fast-deep, fast-shallow). | Only 3 of the 4 regimes present (if fast-shallow is missing, score lower than if missing other regimes). | Only 2 of the 4 habitat regimes present (if fast-shallow or slow-shallow are missing, score low). | Dominated by 1 velocity/depth regime (usually slow-deep). |
| SCORE ____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **5. Channel Alteration** | No channelization or dredging present. | Some channelization present, usually in areas of bridge abutments; evidence of past channelization, i.e., dredging, (greater than past 20 yr) may be present, but recent channelization is not present. | New embankments present on both banks; and 40 to 80% of stream reach channelized and disrupted. | Banks shored with gabion or cement; over 80% of the stream reach channelized and disrupted. |
| SCORE ____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **6. Sediment Deposition** | Little or no enlargement of islands or point bars and less than 5% of the bottom affected by sediment deposition. | Some new increase in bar formation, mostly from coarse gravel; 5-30% of the bottom affected; slight deposition in pools. | Moderate deposition of new gravel, coarse sand on old and new bars; 30-50% of the bottom affected; sediment deposits at obstruction, constriction, and bends; moderate deposition of pools prevalent. | Heavy deposits of fine material, increased bar development; more than 50% of the bottom changing frequently; pools almost absent due to substantial sediment deposition. |
| SCORE ____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **7. Frequency of Riffles** | Occurrence of riffles relatively frequent; distance between riffles divided by the width of the stream equals 5 to 7; variety of habitat. | Occurrence of riffles infrequent; distance between riffles divided by the width of the stream equals 7 to 15. | Occasional riffle or bend; bottom contours provide some habitat; distance between riffles divided by the width of the stream is between 15 and 25. | Generally all flat water or shallow riffles; poor habitat; distance between riffles divided by the width of the stream is between ratio >25. |
| SCORE ____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |

**Figure 2-1. (continued)**

| Habitat Parameter | Category | | | |
|---|---|---|---|---|
| | Optimal | Suboptimal | Marginal | Poor |
| **8. Channel Flow Status** | Water reaches base of both lower banks and minimal amount of channel substrate is exposed. | Water fills >75% of the available channel; or <25% of channel substrate is exposed. | Water fills 25-75% of the available channel and/or riffle substrates are mostly exposed. | Very little water in channel and mostly present as standing pools. |
| **SCORE** _____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **9. Condition of Banks** | Banks stable; no evidence of erosion or bank failure. | Moderately stable; infrequent, small areas of erosion mostly healed over. | Moderately unstable; up to 60% of banks in reach have areas of erosion. | Unstable; many eroded areas; "raw" areas frequent along straight sections and bends; on side slopes, 60-100% of bank has erosional scars. |
| **SCORE** _____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **10. Bank Vegetative Protection** | More than 90% of the streambank surfaces covered by vegetation. | 70-90% of the streambank surfaces covered by vegetation. | 50-70% of the streambank surfaces covered by vegetation. | Less than 50% of the streambank surfaces covered by vegetation. |
| **SCORE** _____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **11. Grazing or Other Disruptive Pressure** | Vegetative disruption, through grazing or mowing, minimal or not evident; almost all plants allowed to grow naturally. | Disruption evident but not affecting full plant growth potential to any great extent; more than one-half of the potential plant stubble height remaining. | Disruption obvious; patches of bare soil or closely cropped vegetation common; less than one-half of the potential plant stubble height remaining. | Disruption of streambank vegetation is very high; vegetation has been removed to 2 inches or less in average stubble height. |
| **SCORE** _____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |
| **12. Riparian Vegetative Zone Width (Least Buffered Side)** | Width of riparian zone >18 meters; human activities (i.e., parking lots, roadbeds, clear-cuts, lawns, or crops) have not impacted zone. | Width of riparian zone 12-18 meters; human activities have impacted zone only minimally. | Width of riparian zone 6-12 meters; human activities have impacted zone a great deal. | Width of riparian zone <6 meters: little or no riparian vegetation due to human activities. |
| **SCORE** _____ | 20 19 18 17 16 | 15 14 13 12 11 | 10 9 8 7 6 | 5 4 3 2 1 0 |

**Total Score** _____

For subsampling, individual grid squares were randomly selected, then organisms were removed from each selected grid until the desired subsample number (300 organisms) was reached. Then any large organic material (whole leaves, twigs, algal or macrophyte mats) was rinsed, visually inspected, and discarded. Randomly selected grid squares were completely sorted regardless of whether the number of organisms was greater than that needed for the subsample. For the Ohio study, organisms were removed from selected girds until the 300-organism subsample was reached. For the New York study, a series of grids were chosen to constitute a 100-organism subsample and a 200-organism subsample for each sample. These subsamples were maintained separately for identification and storage, then the data were totaled to create the 300-organism subsample. Specimens for both studies were placed in a pre-labeled sample container containing 70 percent ethanol and shipped to Monticello Ecological Research Station (University of Minnesota, Monticello, Minnesota) for identification.

## 2.2.2 Taxonomy

For the RBPIII assessments, all specimens were identified to the lowest practical level, generally genus or species; RBPII assessments used family-level identifications. Both utilized primarily Merritt and Cummins (1984), Wiederholm (1983), Brinkhurst (1986), and Thorpe and Covich (1991). RBPI assessments consisted of field identifications generally to the family level; some identifications were to order.

### 2.2.3 Counting

For metrics calculated from taxa counts, pupae and adults were not included in the calculations if larvae or nymphs of the same taxon were identified in the sample. For those metrics which use counts of individuals, pupae and adults were included in the calculations. Exceptions are described for the Ohio and New York data in Appendix A and B, respectively.

## 2.3 Data Analysis

The data were analyzed using the multimetric approach advocated by Karr (1986), Ohio EPA (1987a; b), Plafkin et al. (1989), and Barbour et al. (1995). Metrics were calculated using the 300-organism subsamples from the Ohio study. For the New York study, metrics were calculated based on both 100- and 300-organism subsamples at both family-level and genus/species-level taxonomy. Further rationale for each of these study designs is presented in Sections 3 and 4 of this document.

### 2.3.1 Development of Bioassessment Scoring Criteria

*Bioassessment values* derived from each metric are normalized into *bioassessment scores* so that multiple metrics, which yield a wide range of values, can be aggregated. *Scoring criteria* are developed for each class of test sites, stratified by geographic region and stream order, by dividing the metric value range into equal quadrisections ranging from the lowest possible value of a metric (usually zero) to either the maximum value obtained or the 95th percentile. The scoring criteria categories for Ohio were equal quadrisections from the lowest possible value to the maximum obtained. In most cases, the maximum value of a metric was exhibited at regional reference stations or at specific upstream stations. In the New York study, the upper end of the range used was the 95th percentile, which was used to control for outliers.

Using the appropriate scoring criteria table (Sections 3.2.1, 4.2.1), all calculated or enumerated metric *values* were normalized into bioassessment *scores* (0, 2, 4, 6), which were summed for a total bioassessment score. The total bioassessment scores of test sites were then compared to the regional reference sites for each station. The test sites were evaluated on the basis of their percent comparability to the reference values. For two sites, the regional reference site was found to be impaired (e.g., Furnace Brook, New York) or unable to be sampled (flooded) (e.g., Breakneck Creek, Ohio). Therefore, the upstream reference site (station CC1 and CR1, respectively) served as the baseline for comparison. The suitability of both sites for reference were further examined by deriving information from individual metrics and habitat assessment parameters, and the site CC1 was found to be slightly to moderately impaired. CR1 also had a degraded biological condition but was not given a rating. This illustrates the problems which can arise when relying on a single reference site, and therefore that the comparison should, when possible, be made to reference conditions rather than to single reference sites.

Some metrics include data from the reference site in their calculation; these are known as "paired" metrics. For those sites that used the impaired upstream reference sites as a baseline for comparison, paired metrics were not included in the final assessment. When biological scores are summed using paired metrics, the site designated as the reference site receives an automatic score of 6 (the highest score) for each paired metric, which can artificially raise the overall bioassessment score for that site. Therefore, if the reference site is not minimally impaired (i.e., has some degradation as does CR1 and CC1), the site assessment is given a score that indicates better biological condition than it actually has, or would have if compared to a truly minimally impaired site.

In any biological assessment, comparison of total bioassessment scores to reference is but the first step, which is followed by inspection of individual parameters that allow one to identify potential cause-and-effect relationships. The severity of impairment (slight, moderate, etc.) is determined by comparison with minimally impaired conditions. The thresholds for impairment categories are typically some portion of the distribution of the conditions of all sites. For example, the 75th percentile of the range of scores can be considered the cutoff for nonimpairment. To do this correctly, multiple (at least three) reference sites should be used. However, these studies were designed with only an upstream reference site and a regional reference site. Thus, the assignment of narrative impairment categories, in general, is based on those found in Plafkin et al. (1989). However, because the reference sites in New York appeared to have organic enrichment, it was decided that the actual impairment category should be interpreted as one category less than those listed in Plafkin et al. (1989).

### 2.3.2 Metrics

The metrics used in the biological evaluation of sites include eight "individual" metrics and four "paired" metrics (Barbour et al. 1992). The paired metrics are those which compare the test site to the upstream reference site for the initial calculations. The following is a brief description of the metrics and their calculations. It is worth noting that some descriptions indicate what we expect to find for "good" or "bad" situations for these assessments (based on ecoregions or stream orders). However, the metric value is actually scored good or bad as compared to the reference condition or reference site(s).

1. **Taxa Richness.** Taxa richness reflects the health of the community through a measurement of the total number of taxa present. Taxa richness is calculated by counting the total number of distinct taxa identified in the sample. Generally, taxa richness increases as water quality, habitat diversity, and habitat suitability increase.

2. **Hilsenhoff Biotic Index (HBI).** The HBI was developed by Hilsenhoff (1982) to summarize the various tolerances of the benthic arthropod community with a single value; tolerance values range from 0 to 10, with 10 being assigned to those taxa usually detected in the most degraded situations (i.e., the most tolerant taxa). Only those taxa for which the tolerance values were available were included in these calculations. The formula for calculating the HBI is:

$$HBI = \sum \frac{x_i t_i}{n}$$

where $x_i$ = number of individuals within a taxon,

$t_i$ = tolerance value of a taxon, and

$n$ = total number of individuals in the sample.

Following the Plafkin et al. (1989) document, the HBI was modified to assess the total benthic community not just arthropods and regional development of tolerance values for various environmental pollutants, in addition to organic pollution (Hilsenhoff 1982, 1987; New York State Department of Environmental Conservation, Albany, New York, *in litt* 2/27/89; Illinois Environmental Protection Agency, Marion, Illinois, *in litt* 6/25/86; and Huggins and Moffett 1988). The primary sources for tolerance values and functional feeding group designations were regional when possible (New York State Department of Environmental Conservation, Albany, New York, *in litt* 2/27/89) and USEPA (1990, draft report). Those stations with a lower HBI value are interpreted as being in better condition, having a lower abundance of individuals within tolerant taxa than individuals in sensitive taxa.

3. **Scraper Functional Feeding Group to Scrapers plus Filterer Collectors (Scr/[Scr + Fil] x 100).** The relative abundance of scrapers and filterer collectors reflects the riffle/run community foodbase. When compared to a reference site, shifts in the dominance of a particular feeding type indicate that a community is responding to an overabundance of a particular food source. Scrapers generally increase with increased diatom abundance and decrease as filamentous algae and aquatic mosses increase. However, filamentous algae and aquatic mosses provide good attachment sites for filterer collectors, which may then increase in abundance. The organic enrichment often responsible for overabundance of filamentous algae can also provide fine organic particles used by filterers. This metric reflects biotic response to nutrient overenrichment. Higher values are considered to indicate better conditions.

4. **Individuals of Ephemeroptera, Plecoptera, and Trichoptera (EPT) Taxa to EPT Taxa Plus Chironomidae (EPT/[EPT + Chironomidae]).** This ratio is used as an indication of community balance and compares the number of individuals of Ephemeroptera, Plecoptera, and Trichoptera (mayflies, stoneflies, and caddisflies, respectively) to the number of individuals of EPT taxa plus Diptera: Chironomidae (midges). A relatively even distribution of all four groups indicates a good biotic condition, as does substantial representation of the sensitive groups Ephemeroptera, Plecoptera, and Trichoptera. Environmental stress is indicated by a disproportionately high number of the generally tolerant Chironomidae, reflected by lower values of this metric.

5. **Percent Contribution of Dominant Taxon ([number of individuals of dominant taxon/total number of individuals of all taxa in sample] x 100).** The percent contribution of the dominant taxon uses the abundance of the numerically dominant taxon, relative to the rest of the sample, as an indication of community balance. The lowest practical taxonomic level (assumed to be genus or species in most instances) yields a more accurate assessment value for this metric. A community dominated by only a few species would indicate environmental stress; thus, lower values for this metric are taken to reflect better conditions.

6. **EPT Index.** The EPT Index is the total number of distinct taxa within the Ephemeroptera, Plecoptera, and Trichoptera (mayflies, stoneflies, and caddisflies, respectively) and summarizes the taxonomic richness of three groups of insects that are generally considered to be pollution-sensitive. This value increases with improving water quality.

7. **Shredder Functional Feeding Group to the Total Number of Individuals Collected ([Shr/Total] x 100).** The abundance of the shredder functional feeding group relative to all other individuals allows evaluation of potential impairment to the riparian zone. Higher ratios generally indicate better conditions. Shredders should decrease in abundance if their food source is reduced through habitat alterations or contaminated by toxins.

8. **Hydropsychidae to Total Trichoptera ([H/T] x 100).** Though caddisflies (Trichoptera) as a group are usually considered to be pollution-sensitive, a number of taxa within the Hydropsychidae often greatly increase in abundance and density in degraded and organically-enriched waters. This metric is calculated as the number of individuals of Hydropsychidae to the total number of individuals of Trichoptera in the sample. Higher values reflect a dominance of the hydropsychids (low caddisfly diversity), which indicates poorer water quality.

9. **Pinkham-Pearson Community Similarity Index.** This metric measures the degree of similarity in taxonomic composition between the reference sample and the test sample (Pinkham and Pearson 1976). A higher calculated value reflects a higher degree of similarity to the reference sample and presumably better conditions. It is calculated as:

$$PP = \sum \frac{\text{minimum}(x_{ia}, t_{ib})}{\text{maximum}(x_{ia}, t_{ib})}$$

where $x_{ia}$ = number of individuals in the ith species in sample A

and

$x_{ib}$ = number of individuals in the ith species in sample B.

10. **Quantitative Similarity Index - Taxa (QSI-Taxa).** This measure of comparative similarity in taxonomic composition combined with relative abundance between two sampling stations is based on the concept of "percent similarity" (Whittaker 1952; Bray and Curtis 1957). It has been applied by Shackleford (1988) in Arkansas streams and by others in several individual studies in the mid-Atlantic states. It compares two samples in terms of presence/absence of taxa and relative abundances and is calculated as:

$$S_{ab} = \Sigma \min(p_{ia}, p_{ib})$$

where $p_{ia}$ = the relative abundance of species i at station A,

$p_{ib}$ = the relative abundance of species i at station B,

and

$\min(p_{ia}, p_{ib})$ = the minimum value of species i at station A or B in terms of relative abundance.

Relative abundance is the percentage of individuals in the total sample that are of species i. Values for these calculations range from 0 to 100. Samples that are identical have a score of 100; those which have nothing in common have a score of 0. Thus, those test stations which are more similar to selected reference conditions have higher index values and are inferred to have better biological condition.

11. **Dominants in Common - 5 (DIC-5).** The DIC-5 compares the five dominant taxa (as in greatest abundance) between the reference station samples and test station samples. For this metric, the top five taxa (numerically) for each of the two samples are listed. The number of taxa shared in the top five list is the metric value. Values for this metric range from 0 to 5 with 5 being most similar to reference and 0 least similar.

12. **Quantitative Similarity Index - Functional Feeding Group (QSI-FFG).** The QSI-FFG compares the relative abundance of functional feeding groups between two samples with the goal of showing changes in the function of a community. This metric is calculated in the same way as QSI-Taxa except that the numbers of individuals are those within functional feeding groups: filterer collectors, gatherer collectors, shredders, scrapers, miners, predators, and parasites.

## 2.4 Quality Assurance/Quality Control

The quality control elements for the Ohio and New York case studies are provided in Appendix C.